

Forecasting time series with sieve bootstrap

Andrés M. Alonso*, Daniel Peña, Juan Romo 

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 Getafe,
C/Madrid, Spain*

Received 21 February 2000; received in revised form 10 January 2001; accepted 30 January 2001

Abstract

In this paper we propose bootstrap methods for constructing nonparametric prediction intervals for a general class of linear processes. Our approach uses the $AR(\infty)$ -sieve bootstrap procedure based on residual resampling from an autoregressive approximation to the given process. We present a Monte Carlo study comparing the finite sample properties of the sieve bootstrap with those of alternative methods. Finally, we illustrate the performance of the proposed method with a real data example.

MSC: 62M10; 62M20; 62G09

Keywords: Sieve bootstrap; Prediction intervals; Time series; Linear processes

1. Introduction

When studying a time series, one of the main goals is the estimation of forecast intervals based on an observed sample path of the process. The traditional approach of finding prediction intervals for a linear time series assumes that the distribution of the error process is known. Thus, these prediction intervals could be adversely affected by departures from the true underlying distribution. For example, using a Monte Carlo study, Thombs and Schucany (1990) have shown that the standard (Gaussian) Box Jenkins method performs poorly given a skewed bimodal error distribution.

Some bootstrap approaches have been proposed as a distribution free alternative to compute prediction intervals. Stine (1987) proposes a bootstrap method to estimate the prediction mean squared error of the estimated linear predictor of an $AR(p)$ where p is known, assuming that the error distribution is symmetric and with finite moments. Also, for an $AR(p)$ process with known p , and relaxing the assumptions of Stine (1987), Thombs and Schucany (1990) propose a first backward and then forward bootstrap

* Corresponding author.

E-mail address: alonsoam@est.econ.uc3m.es (A.M. Alonso).

method to find the h -steps ahead prediction intervals. Cao et al. (1997) studied a conditional bootstrap method alternative to Thombs and Schucany's proposal, which is computationally much faster. Masarotto (1990) and Grigoletto (1998) propose a bootstrap method for $AR(p)$ processes with finite unknown p , assuming that some consistent estimator \hat{p} is available. Pascual et al. (1998) generalized the conditional bootstrap approach of Cao et al. (1997) to $ARMA(p,q)$ processes with known p and q and they also incorporate the parameter estimation variability.

This paper considers bootstrap methods to construct nonparametric prediction intervals for a more general class of linear processes than those previously studied. The class of linear processes considered can be written as a one-sided infinite-order moving average process

$$X_t - \mu_X = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad t \in \mathbb{Z}, \quad (1)$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of uncorrelated random variables with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2$ and with at most a polynomial decay of the coefficients $\{\psi_j\}_{j=0}^{+\infty}$. This class includes the stationary and invertible $ARMA(p,q)$ processes, since they have an exponential decay of $\{\psi_j\}_{j=0}^{+\infty}$. Our approach uses the $AR(\infty)$ -sieve bootstrap procedure based on residual resampling from a sequence of approximating autoregressive models for $\{X_t\}_{t \in \mathbb{Z}}$ with order $p = p(n)$ that increases as a function of the sample size n . This $AR(\infty)$ bootstrap procedure was first proposed by Kreiss (1988) (see also Kreiss (1992)) and extensions can be found in Bühlmann (1997). Paparoditis and Streitberg (1992) use this approach in a model identification algorithm. This sieve bootstrap has a nice nonparametric property, being model-free within the considered class of linear processes. Thus, the proposed bootstrap prediction intervals could be applied to this more general class of linear models without specifying a finite dimensional model as in previous bootstrap proposals.

The paper is organized as follows. Section 2 introduces the sieve bootstrap for estimating forecast intervals. Section 3 presents a Monte Carlo study comparing the finite sample properties of the sieve bootstrap with those of alternative methods. We show that the average coverage is better when intervals are constructed incorporating the parameter uncertainty, particularly for small sample sizes. Finally, in Section 4 the performance of the proposed method is illustrated with a real data example.

2. Sieve bootstrap forecast intervals

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a real valued, stationary process with expectation $E[X_t] = \mu_X$ that admits a $MA(\infty)$ representation as in (1) with $\sum_{j=0}^{+\infty} \psi_j^2 < \infty$. Under the additional assumption of invertibility we can represent $\{X_t\}_{t \in \mathbb{Z}}$ as a one-sided infinite-order autoregressive process

$$\sum_{j=0}^{+\infty} \phi_j (X_{t-j} - \mu_X) = \varepsilon_t, \quad \phi_0 = 1, \quad t \in \mathbb{Z}, \quad (2)$$

with coefficients $\{\phi_j\}_{j=0}^{+\infty}$ satisfying $\sum_{j=0}^{+\infty} \phi_j^2 < \infty$. This $\text{AR}(\infty)$ representation motivates the sieve bootstrap. The method proceeds as follows:

1. Given a sample $\{X_1, \dots, X_n\}$, select the order $p = p(n)$ of the autoregressive approximation by AICC criterion: $\text{AICC} = -n \log(\sigma^2) + 2(p+1)n/(n-p-2)$, (see Hurvich and Tsai (1989)).

The AICC criterion is a bias-corrected version of AIC (Akaike (1973)), and it has a more extreme penalty for large-order models which counteracts the overfitting nature of AIC. Other order selection criteria such as BIC, Shibata (1980) could be used, but we prefer AICC assuming the view that the true model is complex and not of finite dimension, and also because the AICC is asymptotically efficient for autoregressive models, i.e., it chooses an AR model which achieves the optimal rate of convergence of the mean-square prediction error. Also, as was illustrated by Hurvich and Tsai (1989), the AICC is less affected than AIC or BIC by changes in the value of the maximum order p_{\max} considered. In the Monte Carlo study, we fix $p_{\max} = n/10$ as recommended by Bhansali (1983).

2. Construct some estimators of the autoregressive coefficients: $\hat{\phi}_p = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)^t$. Following Bühlmann (1997) we take the Yule-Walker estimates.
3. Compute the residuals:

$$\hat{\varepsilon}_t = \sum_{j=0}^p \hat{\phi}_j (X_{t-j} - \bar{X}); \quad \hat{\phi}_0 = 1, \quad t \in (p+1, \dots, n). \quad (3)$$

4. Define the empirical distribution function of the centered residuals:

$$\hat{F}_{\hat{\varepsilon}}(x) = (n-p)^{-1} \sum_{t=p+1}^n 1_{\{\hat{\varepsilon}_t \leq x\}}, \quad (4)$$

where $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\varepsilon}^{(\cdot)}$ and $\hat{\varepsilon}^{(\cdot)} = (n-p)^{-1} \sum_{t=p+1}^n \hat{\varepsilon}_t$.

5. Draw a resample ε_t^* of i.i.d. observations from $\hat{F}_{\hat{\varepsilon}}$.
6. Define X_t^* by the recursion:

$$\sum_{j=0}^p \hat{\phi}_j (X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \quad (5)$$

where the starting p observations are equal to \bar{X} .

In practice we generate an $\text{AR}(p)$ resample using (5) with sample size equal to $n+100$ and then discard the first 100 observations. For autoregressive models, other authors fix the first p observations equal to 0 or draw them with equal probability from all the $n-p+1$ possible blocks of consecutive observations of the original series. Asymptotically, the effect of starting values is negligible (cf. Kreiss and Franke (1992)).

Up to this step, the resampling plan coincides with the sieve bootstrap, and is valid for bootstrapping some statistics defined as a functional of a m -dimensional distribution function (see details in Section 3.3 of Bühlmann (1997)). However, it is not effective for bootstrap prediction, because it does not replicate the conditional distribution of

X_{T+h} given the observed data, where T denotes the last observation. But, if we proceed as do Cao et al. (1997) by fixing the last p observations we can obtain resamples of the future values X_{T+h}^* given $X_{T-p+1}^* = X_{T-p+1}, \dots, X_T^* = X_T$.

7. Given $\{X_1^*, \dots, X_T^*\}$ from the previous step, compute the estimation of the autoregressive coefficients: $\hat{\phi}_p^* = (\hat{\phi}_1^*, \dots, \hat{\phi}_p^*)^t$, as in step 2.
8. Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^p \hat{\phi}_j^* (X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \quad (6)$$

where $h > 0$, and $X_t^* = X_t$, for $t \leq T$.

Finally, $F_{X_{T+h}^*}^*(x)$, the bootstrap distribution function of X_{T+h}^* , is used to approximate the unknown distribution of X_{T+h} given the observed sample. As usual, a Monte Carlo estimate $\hat{F}_{X_{T+h}^*}^*(x)$ is obtained by repeating steps 5–8 B times, where $\hat{F}_{X_{T+h}^*}^*(x) = \#\{X_{T+h}^{*,b} \leq x\}/B$, and $b \in \{1, \dots, B\}$. The $(1 - \alpha)\%$ prediction interval for X_{T+h} is given by

$$[Q^*(\alpha/2), Q^*(1 - \alpha/2)], \quad (7)$$

where $Q^*(\cdot) = \hat{F}_{X_{T+h}^*}^{*-1}(\cdot)$ are the quantiles of the estimated bootstrap distribution.

Notice that, if we omit step 7 and use $\hat{\phi}_j$ in recursion (6), our resampling plan is similar though more general to the conditional bootstrap of Cao et al. (1997). Both approaches will be compared in the Monte Carlo study of Section 3. It is possible to modify the previous algorithm in order to incorporate the variability caused by AICC model selection adding a step where we select a new order p' from the resample obtained with recursion (5) and use it in the subsequent steps.

3. Simulation results

First, we compare the sieve bootstrap approaches with the proposal (PRR) of Pascual et al. (1998), and with the standard Gaussian forecast intervals (STG). We use two sieve bootstrap prediction methods: (CS) a conditional sieve, i.e., omitting step 7 in the algorithm of Section 2, and (VS), the complete algorithm. The bootstrap procedure of Pascual et al. (1998) is similar to the algorithm proposed in Section 2, but instead of an approximating AR-sequence it uses the true generating model. We report the results for the following model:

Model 1: $X_t = \varepsilon_t - 0.9\varepsilon_{t-1}$.

Since method PRR uses the correct model, we could interpret their results as benchmarks. In practice, having observed a sample of size n , the model, and particularly p and q , is invariably unknown. In Alonso et al. (2000) we study the performance of PRR method with mild misspecification and we show that this misspecified approach tends to over-estimate the nominal lengths of forecast intervals.

The error distributions F_e considered are the standard normal $\mathcal{N}(0, 1)$, a shifted exponential distribution with zero mean and scale parameter equal to one, and a contaminated distribution $0.9 F_1 + 0.1 F_2$ with $F_1 \sim \mathcal{N}(-1, 1)$ and $F_2 \sim \mathcal{N}(9, 1)$. We take sample sizes $n = 25, 50$, and 100 , leads $h = 1, 2$ and 3 , and nominal coverages $1 - \alpha = 0.8$ and 0.95 .

To compare the different prediction intervals, we use their mean coverage and length, the proportions of observations lying out to the left and to the right of the interval and a combined measure of coverage and length. These quantities are estimated as follows:

1. For a combination of model, sample size and error distribution, simulate a series, and generate $R = 1000$ future values X_{T+h} .
2. For each bootstrap procedure obtain the $(1-\alpha)\%$ prediction interval $[Q_M^*(\alpha/2), Q_M^*(1-\alpha/2)]$ based on $B=1000$ bootstrap resamples (Methods are STG, CS, VS, and PRR).
3. The coverage for each method is estimated as $C_M = \#\{Q_M^*(\alpha/2) \leq X_{T+h}^r \leq Q_M^*(1-\alpha/2)\}/R$, where X_{T+h}^r with $r = 1, \dots, R$, are the R future values generated in first step.

In steps 1 and 2 we obtain the “theoretical” and bootstrap interval lengths using $L_T = X_{T+h}^{[R(1-\alpha/2)]} - X_{T+h}^{[R\alpha/2]}$, and $L_M = Q_M^*(1-\alpha/2) - Q_M^*(\alpha/2)$. Finally, steps 1–3 are repeated $S = 200$ times to obtain $C_{M,i}$, $L_{M,i}$ with $i = 1, \dots, S$, and we calculate the estimates:

$$\begin{aligned}
\bar{C}_M &= S^{-1} \sum C_{M,i} \\
SE(\bar{C}_M) &= (S^{-1}(S-1)^{-1} \sum (C_{M,i} - \bar{C}_M)^2)^{1/2} \\
\bar{L}_M &= S^{-1} \sum L_{M,i} \\
SE(\bar{L}_M) &= (S^{-1}(S-1)^{-1} \sum (L_{M,i} - \bar{L}_M)^2)^{1/2} \\
CQ_M &= |1 - \bar{C}_M/\bar{C}_T| + |1 - \bar{L}_M/\bar{L}_T|,
\end{aligned} \tag{8}$$

where $\bar{L}_T = S^{-1} \sum L_{T,i}$ is the estimated “true” mean interval length, $\bar{C}_T = (1 - \alpha)\%$ is the nominal coverage, and $|\cdot|$ denotes the absolute value. Sometimes, when we compare two methods M_1 and M_2 , the mean coverage \bar{C}_{M_1} is closer to the nominal value than \bar{C}_{M_2} but the corresponding mean length \bar{L}_{M_1} is greater than \bar{L}_{M_2} , and (the important case) greater than \bar{L}_T . The combined measure CQ_M will be used to compare the methods in such situation. The results for Model 1 are presented in Tables 1–3, using the three sample sizes and error distributions, nominal coverage 95%, and lead times $h = 1$ and 3 . The other possible combinations of parameters are available on request to the authors. Essentially, similar results are obtained in all cases.

For Model 1, method PRR have a better performance than CS and VS in terms of mean coverage and length, as expected. Notice that in this case, the sieve approach never uses the correct model. The difference between PRR and VS, since both methods

Table 1
Simulation results for Model 1, with Gaussian errors^a

Lag	Sample size	Method	\bar{C}_M (se)	Cov. (below/above)	\bar{L}_M (se)	CQ_M
h	n	Theoretical	95%	2.50%/2.50%	3.91	0.00
1	25	STG	88.79 (0.63)	5.03/6.17	3.98 (0.05)	0.08
		CS	87.63 (0.74)	5.72/6.65	4.07 (0.06)	0.12
		VS	89.12 (0.65)	5.14/5.75	4.15 (0.06)	0.12
		PRR	91.04 (0.42)	4.44/4.53	4.00 (0.05)	0.06
		STG	90.27 (0.48)	5.30/4.43	3.90 (0.03)	0.06
	50	CS	89.78 (0.57)	5.44/4.78	3.95 (0.05)	0.06
		VS	91.50 (0.48)	4.51/3.98	4.05 (0.04)	0.06
		PRR	91.78 (0.34)	4.18/4.04	3.86 (0.04)	0.06
		STG	92.64 (0.27)	3.53/3.83	3.90 (0.02)	0.04
		CS	92.15 (0.33)	3.69/4.16	3.91 (0.03)	0.04
	100	VS	93.15 (0.28)	3.27/3.59	4.01 (0.03)	0.03
		PRR	94.07 (0.20)	2.93/3.00	4.00 (0.03)	0.02
h	n	Theoretical	95%	2.50%/2.50%	5.28	0.00
3	25	STG	89.89 (0.44)	4.75/5.36	4.76 (0.06)	0.15
		CS	89.75 (0.46)	4.97/5.28	4.79 (0.07)	0.15
		VS	90.61 (0.42)	4.67/4.73	4.86 (0.07)	0.13
		PRR	92.92 (0.31)	3.59/3.49	5.09 (0.06)	0.06
		STG	91.45 (0.34)	4.67/4.68	4.84 (0.04)	0.12
	50	CS	91.40 (0.33)	4.44/4.15	4.90 (0.05)	0.11
		VS	91.90 (0.31)	4.19/3.91	4.95 (0.05)	0.10
		PRR	93.93 (0.23)	3.02/3.04	5.19 (0.04)	0.03
		STG	92.82 (0.21)	3.74/3.84	4.99 (0.03)	0.08
		CS	92.74 (0.23)	3.61/3.65	5.02 (0.03)	0.08
	100	VS	93.13 (0.21)	3.37/3.50	5.07 (0.03)	0.06
		PRR	94.97 (0.14)	2.49/2.55	5.36 (0.03)	0.01

^aNote: Standard error (se) are in parentheses. \bar{C}_M , \bar{L}_M , CQ_M and se's are computed from (8).

incorporate the parameter estimation variability, could be interpreted as a measure of the effect of the finite dimensional approximations involved in sieve bootstrap on the finite sample results.

When comparing the sieve results with STG, we observe that for Gaussian errors and both lead times and for non-Gaussian and $h = 1$, STG outperforms CS. The good performance of the standard normal predictions for skewed distributions was reported by Thombs and Schucany (1990) for finite autoregressions. The complete sieve procedure VS outperforms CS and STG, revealing the importance of parameter estimation variability. Also, as expected, all methods improve coverage with sample size.

As mentioned in the Introduction, the proposed method is applicable to linear models other than ARMA's. In Table 4 we present the results for the following model:

Model 2: X_t is a Gaussian process with autocovariance generating function equal to $G(z) = \sum_{k=-\infty}^{+\infty} \gamma_k z^k$, where $\gamma_k = 1/(|k| + 1)^3$.

Model 2 is simulated using the Cholesky decomposition of the autocovariance matrix (cf. Beran (1994)). Note that Model 2 satisfies Assumption A2 of Bühlmann (1997)

Table 2
Simulation results for Model 1, with exponential errors^a

Lag	Sample size	Method	\bar{C}_M (se)	Cov. (below/above)	\bar{L}_M (se)	CQ_M
h	n	Theoretical	95%	2.50%/2.50%	3.67	0.00
1	25	STG	89.34 (0.93)	3.31/7.35	3.90 (0.08)	0.12
		CS	89.01 (1.19)	6.07/4.92	4.37 (0.11)	0.25
		VS	90.61 (1.07)	4.54/4.85	4.42 (0.11)	0.25
		PRR	91.29 (0.96)	4.63/4.08	3.98 (0.09)	0.12
		STG	91.71 (0.50)	1.70/6.59	3.81 (0.06)	0.07
	50	CS	90.11 (1.04)	5.57/4.33	4.00 (0.09)	0.14
		VS	92.22 (0.68)	3.79/3.99	4.07 (0.09)	0.14
		PRR	92.56 (0.56)	3.75/3.69	3.89 (0.07)	0.08
	100	STG	92.04 (0.51)	2.01/5.95	3.81 (0.04)	0.07
		CS	90.72 (0.96)	5.78/3.49	3.87 (0.05)	0.10
		VS	93.01 (0.78)	3.62/3.37	3.99 (0.05)	0.11
		PRR	93.57 (0.44)	3.30/3.13	3.80 (0.06)	0.05
3	25	Theoretical	95%	2.50%/2.50%	5.74	0.00
		STG	88.65 (0.49)	5.08/6.27	4.60 (0.09)	0.26
		CS	89.48 (0.52)	5.71/4.81	5.03 (0.12)	0.18
		VS	89.85 (0.49)	5.54/4.61	5.06 (0.12)	0.17
		PRR	92.34 (0.43)	3.97/3.69	5.62 (0.13)	0.05
	50	STG	89.86 (0.36)	4.72/5.42	4.77 (0.08)	0.23
		CS	90.99 (0.35)	5.04/3.98	5.13 (0.09)	0.15
		VS	91.33 (0.34)	4.88/3.79	5.22 (0.10)	0.13
		PRR	93.19 (0.31)	3.35/3.46	5.68 (0.12)	0.03
	100	STG	91.33 (0.26)	3.98/4.69	4.89 (0.06)	0.18
		CS	92.48 (0.26)	4.11/3.41	5.26 (0.07)	0.10
		VS	92.72 (0.26)	3.92/3.37	5.30 (0.07)	0.10
		PRR	93.32 (0.20)	3.31/3.38	5.67 (0.07)	0.03

^aNote: Standard error (se) are in parentheses. \bar{C}_M , \bar{L}_M , CQ_M and se's are computed from (8).

with $r = 1$. Sieve bootstrap methods perform reasonably well in Model 2 since the mean coverage and length tend to the nominal values as the sample size grows, and also here VS outperforms CS.

4. Real data example

In this section we illustrate the performance of sieve bootstrap procedures in a real data set consisting of series F of Box and Jenkins (1976), which is modelled as an AR(2). We compute the 1-step and multistep ahead forecasts intervals for the last ten available observations by using the sieve bootstrap (VS) and the Box–Jenkins methodology (BJ). The nominal coverage was fixed to 90%. The computations were implemented in Splus and the code is available on request to the authors.

Series F of Box and Jenkins (1976) consists of the yields from 70 consecutive batches of a chemical process. Figs. 1 and 2 show the 1-step and multistep prediction intervals, for the Box–Jenkins approach using Gaussian maximum likelihood estimates for an AR(2) process (see page 239 in Box and Jenkins (1976)). The upper limits of the

Table 3
Simulation results for Model 1, with contaminated errors^a

Lag	Sample size	Method	\bar{C}_M (se)	Cov. (below/above)	\bar{L}_M (se)	CQ_M
h	n	Theoretical	95%	2.50%/2.50%	12.41	0.00
1	25	STG	86.55 (1.48)	7.33/6.12	10.45 (0.38)	0.25
		CS	84.93 (1.70)	8.50/6.57	13.59 (0.56)	0.20
		VS	86.99 (1.53)	6.91/6.10	13.70 (0.55)	0.19
		PRR	87.95 (1.25)	6.55/5.50	13.39 (0.52)	0.15
		STG	92.24 (0.72)	4.22/3.54	10.89 (0.28)	0.14
	50	CS	90.79 (0.89)	4.64/4.57	11.65 (0.44)	0.10
		VS	93.17 (0.53)	3.50/3.33	12.84 (0.47)	0.07
		PRR	93.43 (0.43)	3.50/3.07	12.94 (0.46)	0.06
	100	STG	92.97 (0.65)	3.77/3.26	10.97 (0.20)	0.14
		CS	91.92 (0.80)	4.33/3.75	11.85 (0.34)	0.08
		VS	93.07 (0.56)	3.40/3.52	11.92 (0.33)	0.06
		PRR	93.61 (0.23)	3.22/3.17	12.05 (0.37)	0.04
h	n	Theoretical	95%	2.50%/2.50%	19.75	0.00
3	25	STG	88.61 (0.50)	5.72/5.67	12.35 (0.45)	0.44
		CS	89.73 (0.56)	5.19/5.08	15.17 (0.59)	0.29
		VS	89.94 (0.52)	5.10/4.95	15.08 (0.58)	0.29
		PRR	91.71 (0.47)	4.17/4.12	17.68 (0.71)	0.14
	50	STG	90.77 (0.31)	4.63/4.60	13.48 (0.35)	0.36
		CS	91.76 (0.36)	4.22/4.02	15.92 (0.49)	0.22
		VS	91.78 (0.34)	4.20/4.01	15.79 (0.50)	0.23
		PRR	92.91 (0.34)	3.60/3.49	17.78 (0.57)	0.12
	100	STG	91.77 (0.17)	4.10/4.13	14.25 (0.27)	0.32
		CS	92.97 (0.23)	3.50/3.53	16.77 (0.40)	0.18
		VS	93.00 (0.23)	3.48/3.52	16.93 (0.41)	0.17
		PRR	94.05 (0.22)	3.04/2.92	18.85 (0.44)	0.06

^aNote: Standard error (se) are in parentheses. \bar{C}_M , \bar{L}_M , CQ_M and se's are computed from (8).

Table 4
Simulation results for Model 2^a

Lag	Sample size	Method	\bar{C}_M (se)	Cov. (below/above)	\bar{L}_M (se)	CQ_M
h	n	Theoretical	95%	2.50%/2.50%	3.89	0.00
1	25	CS	89.39 (0.49)	5.71/4.90	3.61 (0.05)	0.13
		VS	89.64 (0.47)	5.61/4.75	3.62 (0.05)	0.12
		VS	91.50 (0.36)	4.09/4.40	3.70 (0.04)	0.08
	50	VS	92.19 (0.31)	3.71/4.10	3.76 (0.04)	0.06
	100	CS	93.24 (0.23)	3.35/3.40	3.82 (0.03)	0.04
		VS	93.50 (0.21)	3.17/3.33	3.84 (0.03)	0.03
h	n	Theoretical	95%	2.50%/2.50%	3.91	0.00
3	25	CS	89.99 (0.46)	5.37/4.65	3.64 (0.05)	0.12
		VS	90.09 (0.44)	5.35/4.56	3.65 (0.05)	0.12
	50	CS	92.07 (0.29)	4.01/3.93	3.76 (0.04)	0.07
		VS	92.23 (0.29)	3.79/3.97	3.79 (0.04)	0.06
	100	CS	93.57 (0.22)	3.24/3.19	3.87 (0.03)	0.03
		VS	93.69 (0.20)	3.11/3.20	3.88 (0.03)	0.02

^aNote: Standard error (se) are in parentheses. \bar{C}_M , \bar{L}_M , CQ_M and se's are computed from (8).

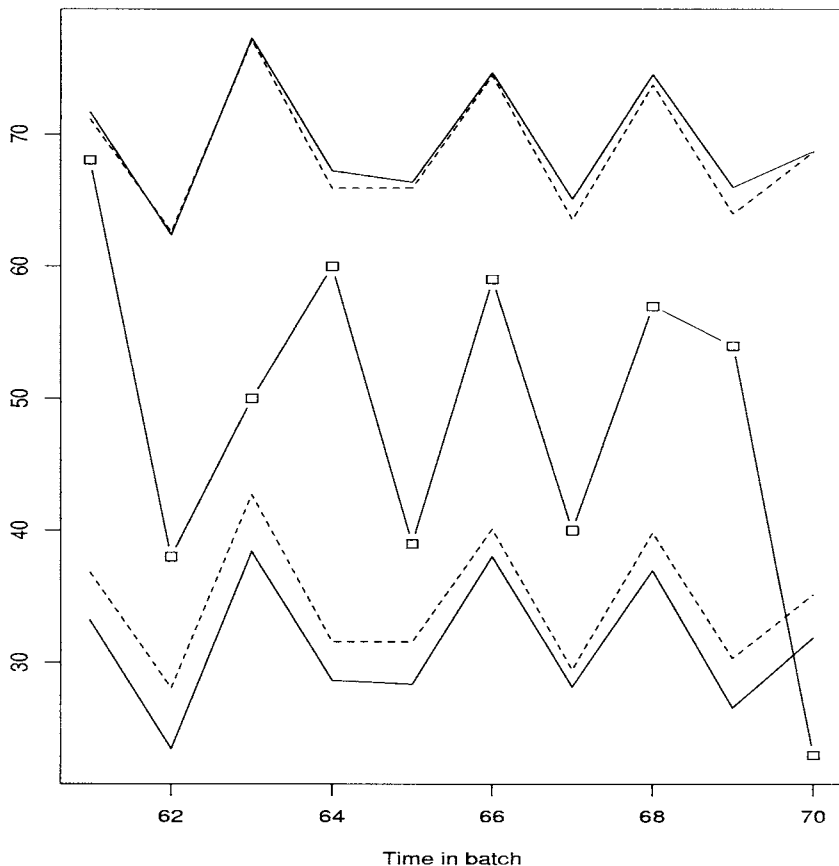


Fig. 1. Observed data (\square), Box Jenkins (---), and VS intervals (—) for 1-step ahead prediction in series F of Box and Jenkins (1976).

intervals are very similar, but in lower limits is observed a downward shift of bootstrap limits revealing the asymmetric distribution of the residuals. The estimated residual skewness is -0.459 with a 95% confidence interval $(-1.130, 0.006)$ and the residual kurtosis is 3.295 with a 95% confidence interval $(2.435, 5.235)$, both constructed by BCA bootstrap method.

5. Conclusions

It has been shown by Thombs and Schucany (1990) that if the order of the AR is known one can obtain prediction intervals by bootstrap with better coverage probabilities than those produced by standard Gaussian procedures. In this paper we have shown that, for general linear models, if we use an AR approximation that grows with the sample size, we can derive a bootstrap procedure for building prediction intervals and

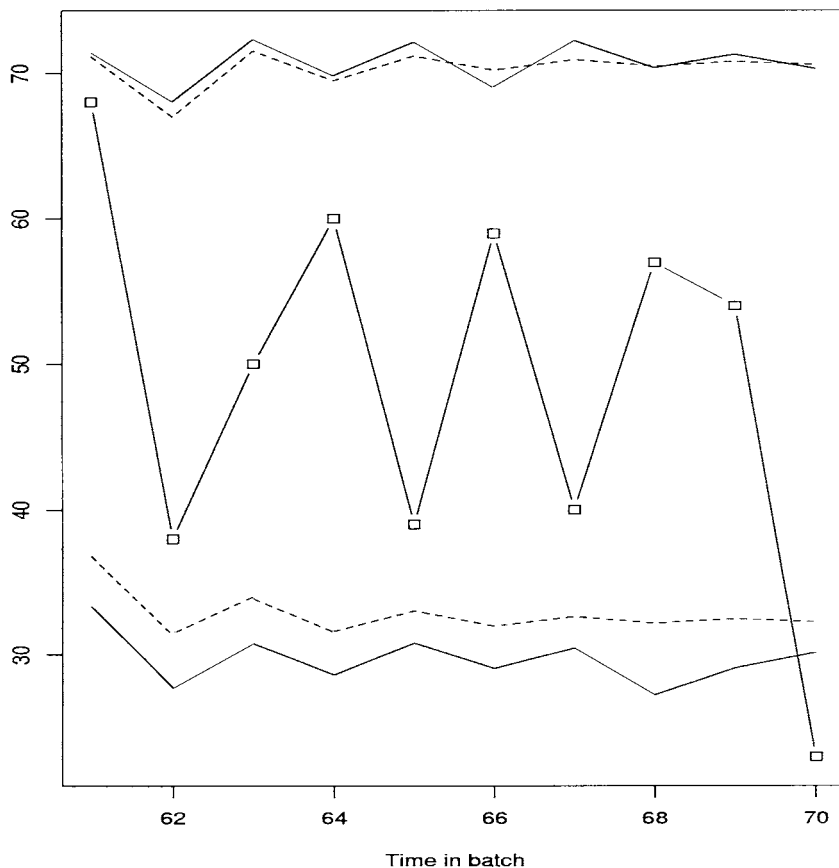


Fig. 2. Observed data (□), Box Jenkins (---), and VS intervals (—) for multistep ahead prediction in series *F* of Box and Jenkins (1976).

Monte Carlo simulations show that it provides better coverage results than previous methods in general cases.

Acknowledgements

We would like to thank Mike Wiper, two referees and the coordinating editor for carefully reading that greatly improved the paper. This research was partially supported by the Dirección General de Educación Superior project DGES PB96-0111 and Cátedra de Calidad BBVA.

References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

- Alonso, A.M., Peña, D., Romo, J., 2000. Forecasting time series with sieve bootstrap. W.P. 00-07, Universidad Carlos III de Madrid, Madrid.
- Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, New York.
- Bhansali, R.J., 1983. A simulation study of autoregressive and window estimators of the inverse autocorrelation function. *Appl. Statist.* 32, 141–149.
- Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA.
- Bühlmann, P., 1997. Sieve bootstrap for time series. *Bernoulli* 3, 123–148.
- Cao, R., Febrero-Bande, M., González-Manteiga, W., Prada-Sánchez, J.M., García-Jurado, I., 1997. Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Comm. Statist. Simulation Comput.* 26, 961–978.
- Grigoletto, M., 1998. Bootstrap prediction intervals for autoregressions: some alternatives. *Intern. J. Forecasting* 14, 447–456.
- Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
- Kreiss, J.-P., 1988. Asymptotical inference of stochastic processes, Ph. D. Thesis, Universität Hamburg.
- Kreiss, J.-P., 1992. Bootstrap procedures for $AR(\infty)$ -processes. In: Jckel, K.H., Rothe, G., Sendler, W. (Eds.), *Bootstrapping and Related Techniques*. Springer, Heidelberg, pp. 107–113.
- Kreiss, J.-P., Franke, J., 1992. Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.* 13, 297–317.
- Masarotto, G., 1990. Bootstrap prediction intervals for autoregressions. *Intern. J. Forecasting* 6, 229–239.
- Paparoditis, E., Streitberg, B., 1992. Order identification statistics in stationary autoregressive moving-average: Vector autocorrelations and the bootstrap. *J. Time Ser. Anal.* 13, 415–434.
- Pascual, L., Romo, J., Ruiz, E., 1998. Bootstrap predictive inference for ARIMA processes. W.P. 98-86, Universidad Carlos III de Madrid, Madrid.
- Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear model. *Ann. Statist.* 8, 147–164.
- Stine, R.A., 1987. Estimating properties of autoregressive forecasts. *J. Amer. Statist. Assoc.* 82, 1072–1078.
- Thombs, L.A., Schucany, W.R., 1990. Bootstrap prediction intervals for autoregression. *J. Amer. Statist. Assoc.* 85, 486–492.